

Title: Assessing L2 Speech Production in a Fully Automated Online Language Test: Challenges, Limitations and Potential

Sylvain Coulange, LIDILEM/LIG Université Grenoble Alpes, France

Abstract:

The landscape of second language (L2) speech assessment is undergoing rapid transformation driven by technological advances, though questions remain about how well current approaches align with contemporary understanding of communicative competence and real-world language use needs (Coulange, 2022; de Jong et al., 2025). The democratization of efficient open-source speech technologies, such as Whisper (Radford, 2022), has made sophisticated analysis of accented L2 speech more accessible than ever before. Combined with increasingly affordable server infrastructure and processing power, along with the widespread availability of quality microphones on student devices, the technical barriers to automated speech assessment have significantly diminished. Furthermore, the emergence of large language models and conversational agents has opened new possibilities for interactive language evaluation (Ma et al., 2025; Huang et al., 2022).

These technological developments coincide with the growing integration of speaking assessment modules in online language tests across both low-stakes and high-stakes contexts. However, a critical gap remains in the transparency and clarity of what these assessments actually measure and how they operate, raising important questions about validity and reliability in automated L2 speech evaluation (Evanini & Zechner, 2019; de Jong et al., 2025).

At Université Grenoble Alpes, we are addressing these challenges through the development of a speaking assessment module for SELF¹ (Cervini & Masperi, 2021), an online language placement test widely used in French universities. This presentation will detail our approach to automatically assessing L2 English spontaneous speech within the specific constraints of placement testing: limited assessment time and the need for brief, independent tasks that can effectively gauge speaking proficiency across diverse learner populations.

We will examine the multifaceted challenges inherent in this endeavor: motivating students to speak naturally in front of their computers, designing tasks that elicit speech samples truly representative of learners' current abilities, adapting assessment protocols to accommodate varying speaker profiles and proficiency levels, and determining which speech features provide the most reliable indicators of L2 competence.

Looking beyond current limitations, we will present our planned evolution toward interactive assessment using a purpose-built conversational agent. This innovative approach combines traditional speech assessment measures with L2-adapted reaction strategies, enabling evaluation of learner comprehensibility across diverse conversational contexts—from informal to formal settings, with interlocutors exhibiting varying degrees of familiarity with L2-accented speech. This methodology promises to assess not only production skills but also interaction competencies, offering a more comprehensive and ecologically valid approach to L2 speaking assessment.

Through this work, we aim to contribute to the growing understanding of how automated technologies can enhance the accessibility, reliability, and authenticity of L2 speech assessment while addressing the practical needs of modern language education.

¹ <https://self.univ-grenoble-alpes.fr/>

References:

- Cervini, C. & Masperi, M. (2021). Conceiving a Multilingual Large-scale Placement Test with Formative Orientation: A Case Study at the University of Grenoble Alpes, In : Lanteigne, Coombe & Brown. Challenges in Language Testing around the world. Insights for language test users, ISBN 978-981-33-4231-6. Springer, Singapore
- Coulangue, S. (2023). Computer-aided pronunciation training in 2022: When pedagogy struggles to catch up. Proceedings of the 7th International Conference on English Pronunciation: Issues and Practices, pp.11-22, 2023. [doi:10.5281/ZENODO.8137754](https://doi.org/10.5281/ZENODO.8137754)
- de Jong, N. H., Raaijmakers, S., & Tigelaar, D. (2025). Developing high-quality, practical, and ethical automated L2 speaking assessments. System, 134, 103796. [doi:10.1016/j.system.2025.103796](https://doi.org/10.1016/j.system.2025.103796)
- Evanini, K., & Zechner, K. (2019). Overview of automated speech scoring. In K. Zechner & K. Evanini (Eds), Automated Speaking Assessment (pp. 3–20). [doi:10.4324/9781315165103-1](https://doi.org/10.4324/9781315165103-1)
- Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. Journal of Computer Assisted Learning, 38(1), 237–257. [doi:10.1111/jcal.12610](https://doi.org/10.1111/jcal.12610)
- Ma, R., Qian, M., Tang, S., Bannò, S., Knill, K. M., & Gales, M. J. F. (2025). Assessment of L2 oral proficiency using speech large language models. [doi:10.48550/ARXIV.2505.21148](https://doi.org/10.48550/ARXIV.2505.21148)
- Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust speech recognition via large-scale weak supervision. [doi:10.48550/ARXIV.2212.04356](https://doi.org/10.48550/ARXIV.2212.04356)